WOMEN'S HEALTH TRAINING RESEARCH AND ADVOCACY CENTRE

EPIDEMIOLOGY
Basic Concept, Research and Women's Health

Dr. P. V. Kotecha

WOHTRAC WORKSHOP REPORT

1997 NO. 5

WOMEN'S STUDIES RESEARCH CENTRE
M.S. UNIVERSITY OF BARODA
VADODARA
INDIA

WOMEN'S STUDIES RESEARCH CENTRE

Epidemiology : Basic Concept, Research and Women's Health

17th - 19th March 1997

ORGANIZED BY

Women's Health Training Research and Advocacy Centre
M. S. University, Vadodara
&
Department of Preventive & Social Medicine,
Faculty of Medicine,
VADODARA.

WORKSHOP CO-ORDINATOR

Dr. P. V. Kotecha[1]

Women's Health Training Research and Advocacy Centre
(WOHTRAC)
(FORD FOUNDATION)
The M. S. University of Baroda
Vadodara
India

---

1 Associate Professor, Preventive & Social Medicine, Medical College, Vadodara.

CONTENTS

FOREWORD

A workshop on "Epidemiology: Basic Concepts, Research and Women's Health" was conducted from 17th to 19th March 1997 under the aegis of WOHTRAC, M. S. University of Baroda and the Department of Preventive and Social Medicine, Faculty of Medicine.

The focus was on various areas of epidemiology some hitherto uncharted, ranging from data generation, measurements, analysis, observations, error searching, use of statistics, planning epidemiological studies and so on. The modus operandi adopted were:

1. Lectures followed by discussions
2. Group work exercises

A total of 58 participants from diverse strata like social scientists, medicine, NGOs, health department, railways, Municipal Corporations, industries and academics interested in Women's Health participated. The faculty comprised of senior teachers from Medical Colleges, Social Sciences and even the students having made a mark in the later category.

The specific objective was to impart understanding in application of basic principles of Epidemiology, to impart issue related to study designs and making them conversant with performing of simple basic analysis of epidemiological data.

It gives me immense pleasure in writing a foreword to the presentation of the deliberation of this workshop by Dr. P. V. Kotecha, Associate Professor of Preventive and Social Medicine, Medical College, Baroda. What follows is the presentation of an extremely methodical mind and the ease with which he has made the entire difficult task look so simple. For even a simple mind to follow speaks of high level of cerebration, application and involvement that to vis-a'-vis an area of Women's Health and item of low priority even in the most of advanced societies.

Dr. Kamal J. Pathak
M. D.
Dean,
Faculty of Medicine

BACKGROUND

Women's Studies Research Center (WSRC) at M. S. University of Baroda was initiated in 1990 with the support of the University Grants Commission. It works for generation building and transmission of knowledge related to women through teaching, research, action and documentation activities. The prime objective of WSRC is to help people to understand, recognize and acknowledge multidimensional roles played by women in the society.

Women's Health, Training, Research and Advocacy Center (WOHTRAC) is a project of WSRC initiated in 1996 with financial support from the Ford Foundation. WOHTRAC is a multidisciplinary partnership venture between M. S. University, Baroda - WSRC, the Population Council and the Society for Operational Research & Training (SORT). The core group members of the WOHTRAC are from diverse fields related to social science like, Human Development & Family Studies, Social Work, Sociology, Community Medicine, Nutrition, Population Research, Communication and Non Governmental Organizations. The major goal of WOHTRAC is to strengthen the institutional capacity at the M. S. University in Social Science Research in the areas of Women's Holistic Health.

While studying various dimensions of social science research and discussing in the core group representing various disciplines, the need for basic concept of quantitative science was strongly perceived by many of us and as an outcome of that need the present workshop on "Epidemiology: Basic Concept, Research & Women's Health" emerged.

WOHTRAC, M. S. University of Baroda and the Department of Preventive & Social Medicine, Faculty of Medicine jointly planned the workshop for 3 full days from 17th to 19th March 1997. The general goal of the workshop was to strengthen education, training practice & research in the field of epidemiology applied to Women's Health.

Specific Objectives:

At the end of the workshop the participants should be able to

1. Understand and apply basic principles of epidemiology to the field of research
2. Understand issues related to epidemiological study design
3. Perform basic (simple) analysis of epidemiological data.

The participants for the workshop were invited from teaching institutions of the university, social scientists, NGOs and medical colleges besides the members of WOHTRAC and the core group. Thus, the workshop was expanded to state level.

Dr. P. V. Kotecha, one of the core group member and a teaching faculty of Preventive & Social Medicine Department with his British Training in Epidemiology at London School of Hygiene & Tropical Medicine volunteered to be the workshop Coordinator.

Organization of the workshop:

The workshop was organized over three full days from 17th March '97 to 19th March '97 at the Conference Hall of Narmada Nigam Guest House, Race Course, Baroda. The two major foci in the contents of the workshop were understanding the theory of basic epidemiology and to have a first hand experience in making meaning from the data and planning epidemiological studies in a given situation. Thus, the workshop course was divided into two parts:

- Lecture followed by discussion
- Group work exercises

Major areas of discussion included

1. Introduction to epidemiology: how is the data generated, what makes data gathering difficult., what are the areas where knowledge of epidemiology is essential and where can that be applied.

2. Measurements in epidemiology : Number of indicators used to measure morbidity and mortality. The terminology classification and what do these terms indicate and what are the limitations of each of these terms. How to calculate different rates and ratios and when to use what parameters

3. Overview of epidemiological studies : What are epidemiological Studies, what are the types & merits and demerits of different types of epidemiological studies.

4. Details of observational and analytic types of epidemiological studies and when to use what type of study. What are the procedures for conducting these studies, what are epidemiological parameters that can be obtained from these studies?

5. Possible sources of error whether statistical or systemic and how to reduce these errors in making epidemiological study.

6. Give a first hand experience of interpreting data generated from studies and discussing these in the group to see different views emerging from the source data daily facilitated by the resource persons.

7. Give a first hand experience of planning the epidemiological studies including sample size and methodology for a given hypothesis or situation, including the control of errors, random or systemic, duly facilitated by the resource person.

8. Presentation of the group work by the participants to the full house and discussion of the issues emerging out of it.

Profile of the Participants:

A total of 58 participants from different fields like social sciences, medicine, NGOs, Health Department, Railways, Municipal Corporation, Industries, Academics participated. The nominations were invited from various organizations working in the field of Women's Health and epidemiology research and based upon the response obtained, participants were selected. These included wide range of people from senior level professors of medical colleges to the students in social science field.

17TH MARCH 1997

INAUGURATION FUNCTION

The inauguration function started with Dr. P. V. Kotecha, the workshop coordinator, welcoming the participants. He welcomed the chief guest Dr. R.V. Bhatt, the senior consulting Gynecologist and Ex. Professor & Head of Obstetric & Gynecology department of Medical College & S.S.G. Hospital, Baroda; Prof. Amitaben Verma, Director, WSRC and Dr. Shagufa Kapadia, Principal Investigator, WOHTRAC.

Dr. Shobha Misra and Dr. Sima Nigam offered the prayer to Maa Saraswati. Then flowers were presented to the chief guest Dr. R.V. Bhatt, Prof. Amitaben Verma & Dr. Shagufa Kapadia. Dr. R.V. Bhatt was then introduced by Dr. R.K. Baxi and was requested to inaugurate the workshop by lighting the lamp.

INAUGURAL ADDRESS

COMMUNITY HEALTH VERSUS INDIVIDUAL HEALTH

By Dr. R. V. Bhatt

In all progressive societies, attention is gradually shifting from individual health to community health. The governments of many countries are spending more money to improve the total health of the community rather than treating individual health problems. It is true that if community health is achieved, many of the individual health problems may disappear.

Individual health requires the attention of medical fraternity to a large extent; whereas community health is teamwork between medical fraternity, community medicine and social scientist. The success in community health depends on the degree to which meaningful coordination is achieved between these disciplines. The community health has improved where these disciplines work as a team. In India the team approach is just taking shape. It would be appropriate to realize the role of clinical medicine, community medicine and social scientist so that better and meaningful approach could be planned.

In clinical medicine, the patient who is sick goes to the doctor. In community medicine, the epidemiologist goes to the community to find the source of the disease so that community may be saved from the spread of the disease. The social scientist is a link -an important link between these two disciplines. An average clinician is so much engrossed in purely clinical aspect and has no knowledge or very little knowledge about disease pattern and its implications. Average clinician is blissfully ignorant about 'case control study' or 'Cohort study' or 'Regression analysis'. He has incomplete knowledge about 'sensitivity' or 'specificity' of any laboratory test.

The epidemiologist is not able to establish equitable wavelength for discussion of common medical problems. Epidemiologist is often not found in clinical wards to see and study disease problems. He feels that his duty ends with field studies. He is not interested in individual clinical response. The social scientist has little knowledge about various diseases, the diagnosis and management. Social scientist has his own world of 'Buzz words' and familiar jargons. He is more devoted to 'demographic transition' or 'social impact' or 'impact on society'. He may not have much appreciation and knowledge about clinical aspect of oral pill or intrauterine device. It is possible that some social scientists may not have seen insertion of intra-uterine device or may not have seen technique of MTP. How does one expect a good critical evaluation of these without some knowledge of the clinical aspect?

Therefore there is an urgent need for these disciplines to widen their horizons and take an active interest in understanding the methodology of the other two disciplines. There is a need for clinician to have better knowledge of bio-statistics, and know some details about various methodology of studying disease pattern. There is a need for epidemiologist to have better insight of the clinical aspect of diseases and for social scientists to understand more about the diseases and various clinical methodologies used to treat diseases.

Dr. Bhatt ended his inaugural address by wishing the workshop a grand success and hoped that the requirement of 'team approach' in community health will be appreciated and learning to participate in 'Inter disciplinary' approach to community health will be encouraged.

Dr. Amita Verma, the Director, WSRC then briefed the history of how WSRC came into existence and described its past and current activities in brief.

Dr. Shagufa Kapadia, the Principal Investigator, WOHTRAC in her very informative talk introduced the activities of WOHTRAC, particularly with reference to research, partnership in research with various faculties, WOHTRAC's role in advocacy for the Women's Health issues in several different ways.

The inaugural function ended with vote of thanks from Dr. Sushma Baxi, Assistant Professor of Obstetric and Gynecology and a participant for the workshop.

The workshop sessions began after the inaugural tea with self-introduction & background of each participant.

INTRODUCTION TO EPIDEMIOLOGY

By Dr. P. V. Kotecha

Dr. Kotecha began the session by giving sets of data and invited comments from the participants. The interpretation among the participants though agreed by majority remained varied. The issues highlighted by Dr. Kotecha were:

- How these information would have been collected
- What must be the study size to generalize it for the whole country
- What are the alternative explanations on the given data and
- Where lies the common fallacies (intended or otherwise) in the interpretation.

Epidemiology differs from clinical medicine in that unit of the interest is the population and not individual. Epidemiology studies the distribution, frequency and determinants of health problems and disease in human populations. The purpose of epidemiology is to obtain, interpret and use health information to promote health and reduce disease. Thus it is concerned in broad sense not only to disease but also in control and prevention of it by adequate monitoring and care where it joins its objectives with hygienists.

Epidemiology is a crucial discipline for the promotion of health of the people wherever they are. It provides a set of skills, approaches and philosophy. That allows causes of health problems to be detected, the association between ill health and a variety of risk factors to be quantified, treatments and public health interventions to be tested, and changes in states of health over time to be monitored.

Epidemiology is a discipline, which allows the distribution of health and ill health in a population to be described: what is the problem and its frequency? Who is affected? Where and when does this health problem manifest in the population? Why does it occur in this particular population? It also provides tools for the comparison of the groups of people.

Epidemiology may play a role in identifying the cause of a health problem: is there an association with a risk factor, or can we be sure it is a cause of the disease or health state in question? If the cause is identified, it could be genetic, an infectious agent, a particular chemical, a behavior or some other causative factor. The epidemiological approach provides a framework within which "causation" can be hypothesized and tested.

Once the disease occurs, epidemiology provides a means of monitoring the course and outcome (natural history) of the condition. It also allows us to answer questions regarding the effectiveness of interventions and therapies and their impact on populations. For example, what is efficacy of particular intervention for controlling diseases in communities? Is intervention X more effective than Y? What are the outcomes of those two interventions? Of course, such information needs to be placed alongside other data before a choice of intervention is made: what are the side effects of the treatments? What are the views of consumers and patients about the procedures? How much does the intervention cost? How efficient is it, and what are the implications for equity if the intervention were adopted more widely? Such data help determine where best to allocate the resources.

Identifying health status in the population may facilitate the need for services and the determination of priorities. Given limited resources, public health practitioners are always under pressure to use resources optimally and to produce the greatest return, in the form of health gain, for a given investment of time, money, materials and personnel. Epidemiology cannot do this alone: it needs to interact with health services research, health economics and other social sciences if wise public health decisions are to be made and health promoted on a population level.

Epidemiology thus has many applications and it is essential tool providing useful information about public health problems, their magnitude and distribution, causation, prevention, prognosis and treatment, and likely impact of interventions.

If epidemiological data are to be of use in policy-making, they need to address questions for which policy makers require answers and need to do so in a convincing and compelling way.

Although epidemiology is just one contributor to health policy debates, it has a crucial role in providing vital data upon which appropriate policy may be based. Increasing our understanding of health problems and their determinants and possible solutions places us in a better position to make appropriate policy.

This course will provide you with the tools to understand basic epidemiological terms and concepts, to use epidemiological data, to draw on epidemiological skills when required, and to recognize what you can gain through epidemiological study. It will help you to appreciate the jargons used by epidemiologists, and will assist you to critically evaluate published research, or to commission appropriate studies. It will also help you appreciate the limitations of the scientific techniques we have at our disposal, the necessity for interacting with other disciplines, and the important, but often indirect, contribution of epidemiology to appropriate public health policy.

Dr. Kotecha described the uses of epidemiology for health planners, evaluators, researchers and academicians giving examples of descriptive, analytic & interventional studies.

MEASUREMENTS IN EPIDEMIOLOGY

By Dr. V.K. Desai

Epidemiology primarily being a quantitative science, measurement is a very important component. Large number of epidemiological parameters used from basic epidemiology to advanced one are essentially measurement of morbidity & mortality measured as rates or ratios. Dr. Desai described these rates in simple terminology explaining the uses and limitations of them. Few basic and most widely used terminology that are often mistakenly used as alternatives of each other were explained.

Prevalence (Point Prevalence)

= Total number of cases of a defined condition existing at a specified point in time, in a defined population.

Incidence

= Total number of new cases of a defined condition, which occur during a specified period in time in a defined population.

In an epidemiologically stable situation (with constant incidence & prevalence over time)

Incidence x Average Duration = Prevalence

Rates: For comparison, incidence and prevalence are not appropriate as they are influenced by the size of the population of the area under study. Rates are appropriate for comparison.

$$\text{(Point) Prevalence Rate} = \frac{\text{Total number of cases of a defined condition existing at a Specified point in time}}{\text{Population at risk}}$$

$$\text{Incidence Rate} = \frac{\text{Total number of new cases of a defined condition which Occur during a specified period of time}}{\text{Population at risk}}$$

Important Notes in Measurements:

1. Clear definitions of both numerator (affected individual) and denominator (population at risk) population are essential.
2. Numerator & denominator should be measured at/over the same point / period in time.
3. When comparing rates, ensure that "***factors***" (i.e. per 100 or 1000 etc.) are the same for all the population. Compared incidence rates should be expressed in terms of similar time duration.
4. All individuals or events in numerator should be included in the denominator.
5. All individuals or events in denominator should be "***at risk***" of moving into numerator (exception: some crude rates: Birth Rates, U5MR).

Prevalence is strongly affected by:

a) Incidence

b) Duration of illness (and survival)

c) Immigration/emigration of cases

d) Loss to follow-up

e) Changes in case detection

f) Changes in population size

OVERVIEW OF EPIDEMIOLOGICAL STUDY DESIGN :

By Dr. B.S. Bhavsar

Fig. 1: Types of Epidemiological Studies

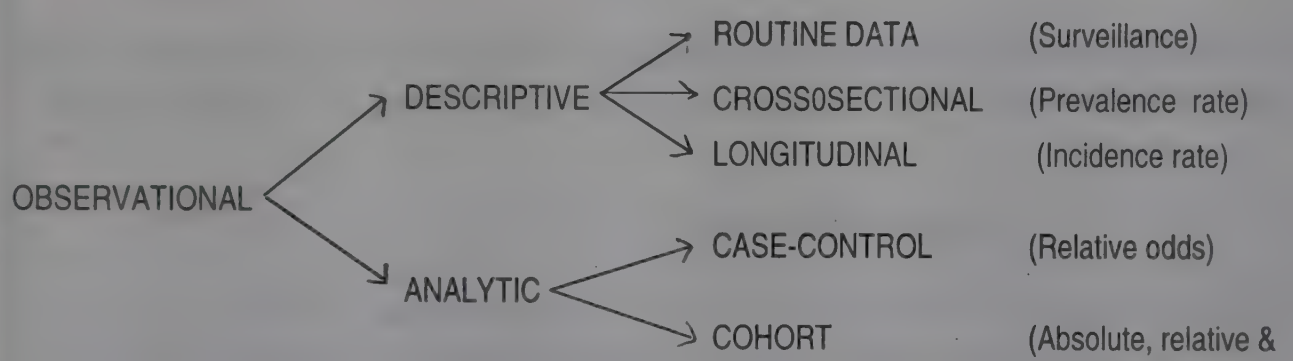


Fig. 1 describes the types of epidemiological studies usually carried out to identify or evaluate public health problems & interventions.

The purpose of epidemiological study needs to be clear. Depending upon the purpose and resources available, the design of the study is selected.

I. Observational

Observational studies emphasize the passive inspection in the process being studied.

A **Descriptive studies** aim at a description of the amount (frequency) and pattern (distribution) of a disease in a population. They answer the questions "What?", "Why?", "When?", "Where?".

Hypothesis testing and controls need not be involved :

1. Routine Date studies utilize routinely available data such as data supporting, diseases notifications, hospital records, registries etc. in order to monitor disease tragedy and suggest explanatory hypothesis.

2. Cross sectional : Studies entail the observation of individuals. They collect prevalence information.

3. Longitudinal Studies entail the observation of individuals over a period of time, allowing measures of incidence.

B **Analytical studies** are essentially comparative, based either on the comparison of the past experience between diseased and non-diseased individuals, or else on the comparison of subsequent disease incidence between individuals with and without some attribute or experience. They are appropriate to test hypothesis as how or why a disease process occurs and to measure the magnitude of risk associated with causative or cause related factors.

1. Case control studies involve the comparison of past experience between 'cases' (deceased individuals) and 'controls' (generally individuals without disease) in search of association between the disease and possible causative or 'risk' factors.

2. Cohort studies involve the comparison of incidence (or recovery) rates between two or more groups of individuals who share different experiences or "Risk factors", in order to assess the implications of these different experiences or attributes.

II **Intervention**

Intervention studies are associated with an active effort to intervene in a disease process in (a segment of) a population, and attempt to access the impact of that intervention on incidence to recovery rates. Their logic is similar to that of the analytic cohort studies.

SMALL GROUP-WORK

The participants were divided to four small groups and medical and non-medical participants were present in all the groups. Each group had a chairperson and a reporter duly selected by the group. Generally a non-medical participant was encouraged to accept the responsibility to report group work to the full house. The exercises given focused on the measurement in epidemiology and had two purposes:

- To have a first hand experience on deriving measurement parameters from the data.
- To interpret the measurement data with their strength & weakness.

Ex. 1 A village having 4000 population had no medical facilities and a cross sectional data in July suggested following prevalence rate of the diseases:

50 malaria/1000 population

10 diarrhoea/1000 population

25 diabetes/1000 population

40 hypertension/1000 population.

The conditions remaining same otherwise, a good doctor starts practicing there, what do you expect the prevalence rate of each of these diseases to show increase, decrease or no change? Explain your answer.

Discussion:

Group discussed the factors affecting prevalence of the disease and distinctly understood the difference between point prevalence and period prevalence like in the given example July being "season" for transmission of both diarrhea and malaria, does not really

reflect the situation round the year. An area not having medical facility previously would show "decline" in prevalence of the disease logically, may in reality be quite contrary at least for a period of time because (a) more reporting for treatments (b) more diagnosis (skill) (c) better record keeping etc. Similarly the group perceived that actual prevalence can drop in a reasonable time only for the infectious, while disease like diabetes and hypertension may not change or even increase (better treatment better survival of patients!)

Ex. 2 In two different anganwadi A & B following workoutput were registered.

	A	B
T.T.1	143	190
T.T.2	140	180
Polio O	110	170
DPT1-Polio 1	100	160
DPT2-Polio 2	105	160
DPT3-Polio 3	105	160
Measles	98	150

You were to report the performance of anganwadi & give your suggestion(s). How would you do that?

Discussion

The discussion began with more work output seen in Anganwadi B as a better Anganwadi. However at a second thought on thinking little more revealed that what is the population covered by Anganwadi and how much it has achieved from what it needs to have achieved is the real way to measure or compare the performance. Thus it was a revealing experience for the participants to critically review the given data and not be guided by what is apparent at the first sight.

Both the exercises were discussed in the group at length and then presented with the help of overhead projector to the entire gathering of the participants.

Ex. 3 In a village having 5000 population 2400 were female and 2600 were male. Following data were noted for the year 1996.

1996	M	F
Jan	10	12
Feb	20	18
March	12	14
April	18	20
May	18	20
June	20	20
July	25	23
Aug	30	32
Sep	23	19
Oct	12	14
Nov	10	12
Dec	10	12

Q.1 What is the incidence rate of malaria for male and female in this village?

Q.2 When all the male and female were asked on 31st dec. whether they had ever suffered from malaria in the entire year of 1996, 150 males and 120 females said yes. Does this finding tally with the registered data? How would you explain.

Q.3 Would you like to express your findings in more than one way?

Q.4 What more information would have interested you?

Discussion

The exercise given to the group recorded cases of malaria month wise in a village population. Incidence rate was calculated for both male and female. Regarding annual total not matching with the annual figure arrived at by "recall" discussion brought out several possible reasons for this mismatch e.g. more than one episode suffered, poor recall, undiagnosed cases, misdiagnosed cases etc. Data when asked to be presented in a different way group thought of putting it as rate common to entire population, devide it quarterly to relate with seasonal transmission. Group wanted information on deaths, deaths due to malaria, diagnosis of malaria, species specific death etc.

18th March 1997

CASE CONTROL STUDIES - WHAT SHOULD WE KNOW?

By Maj. (Dr.) A.R.N. Setalvad

What Is A Case Control Study?

Case control studies are relatively simple and economical to carry out and are increasingly used to investigate causes of diseases, especially rare diseases. The term incorporates a case [i.e. people with a disease or other outcome variable of interest] and a suitable control group [i.e. comparison or reference group of people unaffected by the disease or outcome variable]. In the elementary form of this study, the starting point is the affected person {say patient of carcinoma of lung} and the investigation lies in uncovering the features in his history which may have led to the condition {say cigarette smoking, occupational exposure, air pollution, etc.}. Does one or more of these features appear more frequently in the histories of the affected person than in the histories of the unaffected persons? This is the answer we seek from such a study by comparing the relative frequencies between the "cases" and "controls". The occurrence of the possible cause is compared between cases and controls. Data concerning more than one point in time are collected. Case control studies are thus in essence longitudinal, in contrast to cross-sectional studies. They are also called retrospective studies since the investigator is looking backwards from the disease to a possible cause. However this is a misnomer as case control study can also be prospective in which data collection continues with the passage of time or the exposure data are collected before the development of the disease. A case control study may as well as be a prospective inquiry.

Selection of Controls & Cases

A case control study begins with the selection of cases, which should represent all the cases from a specified population {Fig. 1}. The most difficult task is to select controls so as to sample the exposure prevalence in the population that generated the cases. Furthermore, the choice of controls and cases must not be influenced by exposure status, which should be determined in the same manner for both. It is not necessary for cases and controls to be all-inclusive; in fact they can be restricted to any specified subgroup such as old people, males or females, etc. While it is a matter of flexibility dependent upon the relative availability of cases, what is important is to ensure that there is sufficient similarity between cases and controls when the data are to be analyzed. e.g. if the

data were to be the older age groups, the study would be inefficient leading to misleading conclusions and much of the time and effort would be wasted. The controls should represent people who would have been designated study cases if they had developed the disease. Ideally, case control studies use new cases (i.e. incidence) to avoid the difficulty of disentangling factors related to causation and survival; in diseases like congenital malformations, prevalence data can also be used.

Start & Duration of Exposure

An important aspect of case control studies is the determination of the start and duration of exposure for cases and controls. In a typical design, the exposure status of the cases is usually determined after the development of the disease, and usually by direct questioning of the affected person or a relative or a friend. This may lead to what is know as "Recall Bias" (vide infra).

Precision of Study

The precision of a value calculated from a study depends upon the two factors:

(a) Size of sample; and

(b) Variability of the characteristic within the population from which the sample is drawn.

The statistician's aim is to pass from these simple rule to more precise formulae which will enable him to estimate with a certain degree of confidence the value of occurrence, etc. {i.e. proportion} in the population at large. He further aims to avoid erroneous conclusions from differences between proportions etc. when these differences could easily have arisen by chance. The precision of a study can also be improved by ensuring that the groups are of appropriate relative size. This is often an issue in a case control study when a decision is required on the number of controls to be chosen for each case. Because of its relative speed and cheapness as compared with other approaches, case control studies is a preferred method; there are difficulties in selection of cases, selection of controls and obtaining the data once the selections are made. If these precautions are not observed, case control studies often produce contradictory and conflicting results which are wrongly attributed to small or inadequate sample size.

A problem that often crops up while determining the size of sample is the number of control as related with the case. The first problem in a case-control study is the selection of cases, which usually receives little consideration beyond a definition of the type of the diseases and a statement about the confirmation of the diagnosis. However, the patients do not exist in isolation; the disease is the end result of a process that is almost always complex and not easy to quantify. For greater difficult is the selection of the controls. While there cannot be a rule of thumb covering all situations, there is little sense in having more than four controls for each case; optimal figure being a matching control for each case or two controls for each case. There are two sets of controls: general population and group of patients having other diseases. Although these two populations are not clearly identical, the latter are usually preferred due to ease of availability.

Selection Bias

The quantification of selection bias is often difficult; matching for some variables does not ensure comparability. If individuals entering or remaining in a study display different associations from those who do not, usually a biased estimate of the association between exposure and outcome is produced.

Measurement or Classification Bias

This occurs when the measurement is inaccurate or partially accurate. There are many sources and their effects are of varying degrees. This can be offset partly by randomizing the analysis patterns. For example, if the specimens of the study and control

groups are analyzed randomly by different laboratories, the errors will be random and potentially less harmful. If measurement bias occurs equally in the groups being compared it almost always results in an underestimate of the true strength of the relationship. This form of bias may also account for some apparent discrepancies between the results of different epidemiological studies.

Recall & Questionnaire Bias

Recall bias is a variant of measurement bias of particular importance in retrospective case control studies. It occurs when there is a differential recall of information by cases and control. For example, the cases may be more likely to recall past exposure especially if it is a well known cause and effect relationship like lack of exercise and heart disease or sickness during pregnancy and birth of a handicapped child. When the data are collected by direct questioning, the informant's answers may be influenced by knowledge about the hypotheses under investigation or the disease experience itself. This can either exaggerate or underestimate the degree of effect associated with the exposure. We may be forced to rely on the unreliable memory of the subjects. This problem can be avoided if accurate exposure data are available from an established recording system [e.g. employment records in an industry for periodical examination reports, etc.] or the case-control study is carried out prospectively - i.e. the exposure data are collected before the development of the disease.

Questionnaire bias arises when the question is interpreted differently by different subjects. This may be due to:

(a) The manner and the setting in which the question is put;

(b) The question is ambiguous;

(c) The perception of the respondents is at variance with that of the worker as regards the problem/question;

(d) The respondents may not understand the question at all but they would be hesitant to admit it.

It is obvious that we must be very careful in framing the questions so as to avoid these pitfalls.

Sampling in Case Control Studies

Although randomized sample of adequate sample size is the best thing to have, most of the case control studies are not conducted in such a manner due to variety of reasons and practical difficulties, particularly if the disease is rare. Any study on human patients requires ethical approval and the consent of the patients which itself may be difficult to obtain. Getting the cooperation of volunteers as controls is itself a daunting task. Therefore in general, medical studies are usually carried out using controls drawn from populations that are much more restricted than those about which we wish to draw conclusions. We may have to restrict the sample size, we may have to use patients from one hospital instead of several hospitals or all patients, or populations from a small area rather than that of a large region or state or country. We have to rely on special volunteer groups like prisoners, medical students, and nurse as controls. All these factors lead to inadequate or truncated sample size that may not truly represent the entire population. Findings from such a study can only apply to the population from which the sample was drawn. Any conclusion, which we derive for wider populations or countries, depends upon evidence that is non-statistical and often subjective and unspecified. This may let us down and the results in one population may not apply to another. In such circumstances, it is desirable that other workers on other populations should repeat such studies, so that we can sample the larger population at least to some extent by pooling the results appropriately.

Cause & Effect Relationship

One of the most difficult tasks in medicine is to determine the cause of disease so that we may devise methods of prevention. However for achieving this we are working in an area where experiments are often neither possible nor ethical. We must also face

the fact that disease effect and putative causes do not exist in isolation but in a complex interplay of many interrelated factors. We have to ensure that the effect which we are observing is not the result of some other factor acting on both the "cause" and "effect". This is more difficult in a case-control study wherein we are going backwards in time that may lead to erroneous conclusions.

Odds Ratio

The association of an exposure and a disease is measured in a case-control study by calculation of "Odds Ratio". It is the ratio of "odds" [i.e. chance] of exposure among the cases to the odds in favor of exposure amongst the controls. It is similar to the risk ration, particularly if the disease is rare. The formula of Odds Ratio is:

$$O.R. = \frac{[\text{Disease Yes \& Exposure Yes}]/[\text{Disease Yes \& No exposure}]}{[\text{No Disease \& Exposure Yes}]/\text{No Disease \& No Exposure}]}$$

$$= \frac{[\text{Disease Yes \& Exposure Yes}]*[\text{No Disease \& No Exposure}]}{[\text{Disease Yes \& No Exposure}]*[\text{No Disease \& Exposure Yes}]}$$

Examples of Case-Control Studies

A classic example of a case-control study was the discovery of the relation between thalidomide and unusual limb defects in babies born in Germany in 1959-60. The study, undertaken in 1961, compared affected children with normal children. The study produced the following results with the obvious association of thalidomide and congenital malformations:

MALFORMATION	THALIDOMIDE INGESTION DURING 4-9 WEEKS OF GESTATION		TOTAL
	Yes	No	
Yes	41	5	46
No	0	300	300
Total	41	305	346

Another example of a case-control study was the investigation in Papua New Guinea in people with enteritis necroticans with relation to consumption of meat. The study produced the following results, which proved that the disease was more likely to be associated with consumption of meat:

Enteritis Necroticans	Recent meat ingestion		Total
	Yes	No	
Yes	50	11	61
No	16	41	57
Total	66	52	118

In these examples given above, the odds ratio is as under:

[a] Thalidomide study:
O.R. = {41/5}/{0/300}
= {41*300}/{5*0}
= 12300/0
= ∞ (Infinite)

[b] Enteritis study:
O.R. = {50/11}/{16\41}
= {50*41}/{11*16}
= 2050/176
= 11.6

With the obvious conclusion! Which means that the cases were 11.6 time more likely than controls to have ingested recent meat.

Utility of Case-Control Studies

The relative utility of a case-control study is summarized below:

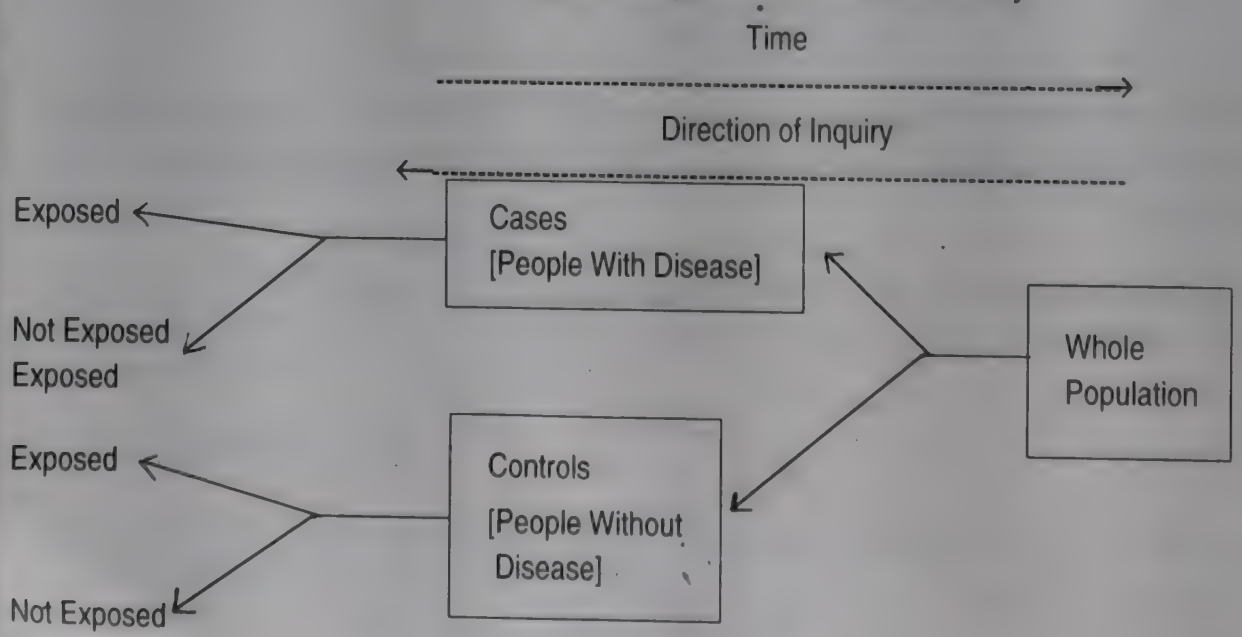
Type of Study	Relative Utility
Investigation of rare disease	+++++
Investigation of rare cause	-
Testing multiple effects of cause	-
Study of multiple exposure and determinants	++++
Measurement of time relationship	+ @
Direct measurement of incidence	+ $
Investigation of long latent periods	+++

Note: +++++ indicates degree of suitability
- Indicates not suitable
@ If prospective
$ If population based

Conclusion

An attempt has been made to narrate some of the practical problems associated with case control studies. If these factors are kept in mind while designing the study itself, many of the inherent problems of a case control study can be avoided to a great extent. One advantage of a case control study is that it is likely to give the answer more speedily than the cohort study and in some rare diseases, perhaps it is likely to be the only possible approach. Although case control studies are subject to several forms of bias, the results from large well designed investigations of this kind provide good evidence for the causal nature of an association; judgments often have to be made in the absence of data from other types of study.

Fig. 1 Design of a Case-Control Study



INAUGURAL FUNCTION

CHIEF GUEST : DR. R.V. BHATT

WORKSHOP CO-ORDINATOR : DR. P.V. KOTECHA

TECHNICAL SESSION IN PROGRESS

SESSIONS ENDED BUT DISCUSSIONS DID NOT

COHORT STUDIES

By Dr. Dilip Mavlankar

After briefly describing the history of origin of epidemiology study (story of John Snow and the Cholera epidemic in London City). Dr. Mavlankar described cohort study.

Cohort studies involve the comparison of incidence (or recovery) rates between two or more groups of individuals who share different experiences or attributes in order to access the implications of these different experiences or attributes.

Fig.

		Disease +	Disease –	
Study Begins --	Factor +			1000
	Factor _			1000
		Disease +	Disease –	
Study Ends --	Factor +	250	750	1000
	_	100	900	1000
		300	1650	2000

1. The group of persons to be studied (the cohorts) are defined in terms of characteristics manifest prior to the appearance of disease under investigations.
2. The study groups or defined and observed over a period of time (i.e. longitudinally) in order to determine the incidence rates of the disease/event among them.

Fig. 2 Relative Utility Of Cohort Studies

The relative utility of a cohort study is summarized below:

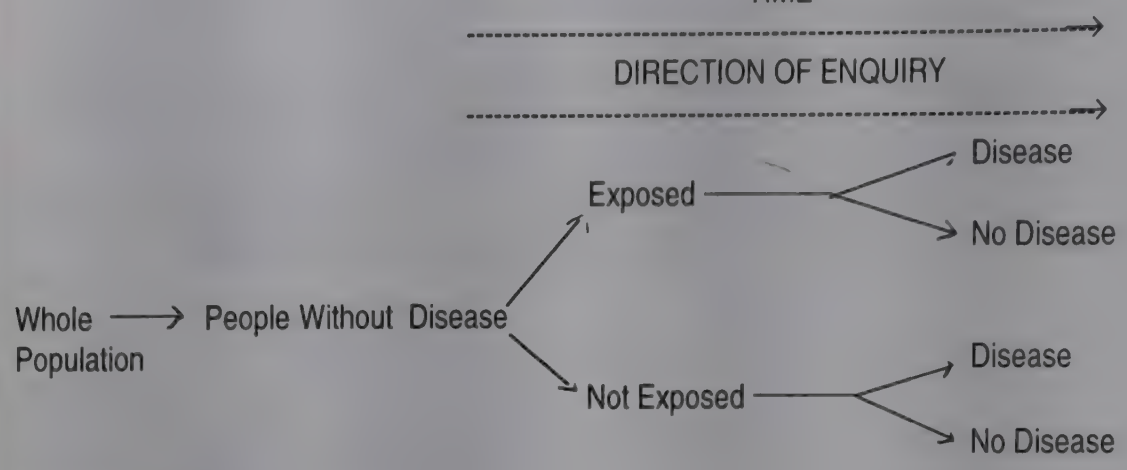
Type of Study	Relative Utility
Investigation of rare disease	–
Investigation of rare cause	+++++
Testing multiple effects of cause	+++++
Study of multiple exposures and determinants	+++
Measurement of time relationship	+++++
Direct measurement of incidence	+++++
Investigation of long latent periods	-

Note : + ... +++++ indicates degree of suitability indicates not suitable.

Relevant Points of Cohort Studies

1. Generally prospective, thus organization difficult and costly in time & personal.
2. Sampling consideration: Representatives, selection, numbers, varying exposure.
3. Comparability cohorts should be similar except for attribute differences. Standardization technique.
4. Consistent methodology is essential.

Fig. 3 Design of a Cohort Study



Follow-up

Should be through, and equal for different cohort losses, to follow-up examine for bias. Unequal follow-up times "person-years" denominator.

Results of Cohort Study

a) Absolute Risks

Incidence rates in expose = R (+)

Incidence rates in unexposed = R (-)

b) Relative Risk = $\frac{\text{Incidence rates in "Exposed" R (+)}}{\text{Incidence rates in "Unexposed" R (-)}}$

Relative risk is a useful measure of strength of an association with a risk factor. "Causal" risk factors associated with high relative risk.

ERRORS IN EPIDEMIOLOGICAL STUDY:

By Dr. P. V. Kotecha

Epidemiology study is a measurement exercise. One needs to be clear what is one measuring. The more clarity one has better is the planning for the study and less are the errors. Forming question correctly helps planning study better.

"Does this iron therapy reduce anemia prevalence?" is a simple question. This is better redefined as "How much reduction of anemia prevalence this iron therapy will make if used for x months when initially anemia is prevalent; in the defined population of defined age group?"

What do we measure, do we have clarity of selection criteria, exclusion criteria and how accurate and valid are our measurements?

When ever there are measurements there are possibilities of error. In other words the measurement is subjected to errors which are broadly classified into two types:

1 Random Error

2 Systemic Error

Random Error :

Results in an estimate being equally likely to be above or below the true values. Major component of random error is sampling error. Random error is expressed as

I) type 1 error (also called alpha error)

Numerically this coincides with the traditional p-values. This indicates the level of error/probability of findings of observation which in true sense does not exist.

ii) type 2 error (also called Beta error)

This is the level of error/probability that indicates the probability of finding no difference when travels it exists.

Null Hypothesis: For statistics work interpretation initially we spell risk factor and outcome. This is called Null Hypothesis.

	Null Hypothesis	
	Real Situation	
	True	False
study results	True: True-ok False: Type II - error	False: Type I - errors True: true-ok

Random error: can be reduced by

- Increase in Sample Size
- Matching

Systemic Error (also called Bias) :

Any effect at any stage of investigations or inference tending to produce results that departs systemically from true values.

Bias thus is preferentially and/or predictably in one director or other. Thus it will either under estimate or over estimate the measurement.

There are 3 types of biases:

1. Selection Bias
2. Information Bias
3. Confounders.

Selection Bias: Relationship between outcome and exposure is different in those who actually participate and those who would be theoretically eligible for study but do not participate. Thus the bias is at selection level.

Information Bias: Once the population is fixed or determined information about being diseased or not diseased OR being exposed to a factor or not exposed is to be collected. Bias here is called information bias.

- Differential misclassification
- Non differential misclassification

Cofounders : A factor that discloses the apparent magnitude of the effects of a study factor or risk. Such factor is **determinant** of the outcome of interest and is **UNEQUALLY** distributed among exposed & not exposed.

	Anemia	No Anemia	Total
Hindu	360 (36%)	640	1000
Muslim	450 (45%)	550	1000
Total	810	1190	2000

Muslim ladies have 9% more anemia then Hindu ladies.

But then:

	Anemia in Hindu Women	Anemia in Muslim Women
High SES	240/800 (30%)	150/500 (30%)
Low SES	120/200 (60%)	300/500 (60%)
Total	360/1000 (36%)	450/1000 (45%)

1) SES is *related* to anemia

2) SES is present in both the groups in *different* proportion

Therefore it is a confounder,

How to Control the Confounder?

Design Level	Analysis Level
- Randomize	- Stratification
- Restriction	- Statistical Modeling
- Mortality	

Summarizing about controlling errors in epidemiological Studies Dr. Kotecha emphasized.

1. Good Study Design
2. Proper Sampling/Selection Procedures
3. Standardization of procedures adopted in data collection
4. Training & retraining the staff
5. Pre-testing & piloting
6. Blinding of the study
7. Continuous Quality Control
8. Data clearing (GIGO principle!)
9. Appropriate Analysis
10. Quantify Errors & accept them
11. Moderate Interpretation

- Common Sense & logic ONLY point to Bias.

Use Grey matters of the brain generously and judiciously!

SMALL GROUP-WORK

Ex. 4 In an attempt to measure the effect of birth weight on subsequent growth of children a cohort study is carried out. 300 children with birth weight less than 2.5 Kg. were followed till age 1, when anthropometric measurements were made to assess nutritioinal status. A similar number of children born during the same period with birth weight more than 2.5 kg. were also followed up.

The Result is as follows:	<2.5 kg.	>2.5 kg.
No. of children studied	300	300
No. of children detected to be malnourished at age 1	102	51

Q.1 What is the exposure factor that is being studied?
Q.2 What is the occurrence rate of malnutrition among exposed?
Q.3 What is the occurrence rate of malnutrition among not exposed?
Q.4 Calculate the RelativeRisk. Describe in one sentence what does that mean?
Q.5 Calculate the attributable risk an dattributable risk % among exposed. Describe in simple language what does that mean?

Discussion

The exercise pertained to understanding the concept of risk factor or the exposure factor. Few of the participants, participating in this group were not quite clear about "expose" and "risk" as generally referred to in an epidemiological exercise. Then there after, it was easy for the group to calculate rates asked. Explanation of Relative Risk also generated discussion. "Relative to what?" was the most common question and, in answering the same in one line the concept was clarified. Concept of attributable risk and attributable risk percentage was also debated. Exposure factors' contribution to the unfavorable outcome and effect of removal of the same was understood threadbare.

Ex. 5 Following is the result :

	LOW BIRTH WEIGHT (LBW)		NORMAL BIRTH WEIGHT	
S.E.S.	Studied	Mal. at age 1	Studied	Mal. at age 1
Low	180	81	100	30
Middle	80	18	100	15
High	40	3	100	6
Total	300	102	300	51

(A) Calculate the malnutrition rate % for each of the SES categories for the two study groups.
(B) Estimate category specific relative risk.

Discussion

In the example relating birth weight and socioeconomic class. Groups initially had difficulty relating the two tables showing data of distribution of LBW in SE class categories and their follow up. The data in table one was, interpreted as showing equal distribution in all three SE Classes while percentage distribution showed highest LBW among low SC class.

In the data of malnutrition in both groups at age 1, initially only % was taken out showing marked difference among two groups in each SE class. However, R.R. was calculated for each SE class as questioned in the exercise. This simply classified that RR was almost similar in each SE class and what was gross difference only on % distribution, disappeared. This showed higher R.R. for developing malnutrition among normal birth weight babies in the lower SE Class. Comparing this with the fact that R.R. for malnutrition in the same SE class in both groups was comparable, clearly showing that being LBW is NOT the only risk factor which would lead to higher malnutrition. While lower SE class is a factor that contribute to malnutrition slightly more. The entire exercise was found to the quite tough and confusing to begin with and most enlightening at the end. The group got confused, fumbled and at the end understood and enjoyed the process and reasoning of looking at the given data set in more than one ways.

Ex. 6 You are interested in detecting the possible risk factor for pregnant mother which may lead to foetal distress.

Q. What type of epidemiological study will you carry out?

- Form the hypothesis.
- Define the variables.
- Discuss the methodology & possible outcome of the study.
- Enumerate the biases that may have to be considered?
- Ethical Issue?

Discussion

For the exercise on fetal distress in pregnancy one group discussed it as a case control study and they discussed biomedical variables/cofounders which could creep in. They also appreciated the need for uniform definitions for diagnosis of a case and also considered the technical bias that could come from different sets of instruments used by different people.

The other group thought that for the given problem, a cohort study prospective was the best answer. And they also could consider the need for identifying variables/cofounder and need for uniform criteria and the risk of bias creeping into it because of recall/ records instrument, and measurements.

The group after a lively debate considered ethical issues e.g. when a factor is known to the leading to fetal distress(FD) there is NO need to study and allow FD to come up. Similarly keeping the same in mind FD once detected, the need to intervene, on ethical grounds will have to be provided for in the study.

Ex. 7 The relative frequency of newly reported cancers of specific sites in two populations is as follows:

Site of cancer	Percent of Total	
	Population A	Population B
Lung	10.0	6.7
Breast	30.0	20.0
Uterus	25.0	10.7
All other	35.0	56.7
Total all sites	100.0	100.0

Conclusion: Pop. A has higher incidence of Cancer lungs, breast & uterus than B. Give your comments.

Discussion

Exercise on Cancer data has a misleading conclusion. Group quickly agreed to what was concluded and then, reverted to reading the exercise once more. It was realized that data pertains to proportionate comparison without giving the denominator or the actual population in both groups. There are many issues related with comparatively of populations. Those which were discussed, particularly age distribution of the population, as cancer risk would be different at different ages. As some cancers are gender specific, gender segregated population distribution in each age group category should have been given.

Ex. 8 A survey conducted recently in 1996 showed the following relationship between age and alcohol consumption in a population:

Age Group	Percent Abstainers
20-29	25.3
30-39	23.2
40-49	28.4
50-59	47.9
60-69	48.6
70+	68.2
(X^2 test = 49)	($p < .01$)

Conclusion: As persons grow older they tend to give up drinking alcoholic beverages.

Discussion

Alcohol abstinence and age group distribution was an example, which classified the need to look carefully into data showing incorrect association. Here group could think of need to get the population proportion in each age group category because more percent would be misleading. i.e. sample size is mentioned nowhere! Moreover in a percentage distribution we see only the trend. More Alcoholics could have been dead due to alcohol and, removed from the higher age group or perhaps abstainers would have survived longer! A good longitudinal study would have been better than a survey, hence the conclusion drawn is not accepted. Group enjoyed that brain teasing presentation and argued it out well. Main explaination is the cohort effect and not age related.

Ex. 9 In trying out vaccine efficacy of influenza vaccine, 200 volunteers were registered and given vaccine. In 200 controls 30 developed influenza while those who opted for vaccine only 10 developed influenza.

Q.1 What type of epidemiological study this is?

Q.2 Is influanza vaccine effective?

Q.3 Are there any problems with the study?

Q.4 Give your suggestions to improve the study.

Discussion

Participants found the exercise of efficacy of influenza vaccine very interesting and realistic. During the group discussions they considered many issues. Most of them decided it to be field trial as type of experimental study design, however few of them were considering it a case control study as a possibility initially. Regarding efficacy of vaccine most of them derived relative risk and since it came to be 0.33, they were unanimous for their interpretation of it, that is vaccine is protective. Few of them could reach to calculate vaccine efficacy to be 66%.

While discussing problems with the study selection bias, matching was not done (or reported) came out from most participants and few other problems mentioned were; duration of study, adverse effects of vaccine not studied, vaccine strain used as well as that responsible for causation were not studied. In this section lot of limitations of study came out.

Finally after such a brainstorming discussion for almost an hour, participants were very enthusiastic to give their suggestions. They included Random sampling, Blinding (Double/Triple), study adverse effects along with that and need to do the study on a larger scale as well as utility and cost effectiveness of which were discussed.

Ex.10 A congenital abnormality may result in child if mother is exposed to infection during pregnancy. You want confirm whether it is true or not. Plan & discuss the outline of epidemiological study for the above purpose.

Discussion

In this exercise opinion(s) of the group were divided hence they considered chort as well as case control study design.

One sub-group designed a prospective cohort study with the hypothesis that infection during pregnancy is not associated with congenital malformation in a child. They opted to select 5 villages in Baroda district and with the help of health worker proposed to include cases for two years and follow them up. The villages were selected in order to avoid selection bias instead of institutional antenatal cases. Age group was selected to be 20-35 years. As after 35 years of maternal age probability of congenital malformation increases as such. For this purpose infection will be defined with clinically identifiable symptoms. Medicine taken during pregnancy will be noted as it may be responsible for congenital malformation.

The second group considered the possibility of case control study as it will be easy to carry out because of few case of congenital malformation.

However in the end during panel discussion, doing case control study first and then after reaching to probable risk factor prospective cohort study should be followed was in serious consideration. As a whole the exercise was very thought provoking and was of great help in clarifying concepts.

Ex. 11 Infant Mortality is divided into 2 parts neo-natal mortality & post neonatal mortality.

In India neo-natal mortality accounts for only 30% of total infant mortality while in USA neonatal mortality accounts for 75% infant mortality. Comment.

Discussion:

At the outset first group felt that the services in neonatal care may be initially better in India due to the child being with mother and more often breastfed that U.S.A. However such a difference would not occur because of the reasons so thought. Later with the help of the resource person they grasped that these are proportionate data. Infant mortality in U.S.A. is one tenth of Indian data and thus Neonatal Mortality of 30/1000 live birth and post neonatal mortality of 70/1000 live birth in India and Neonatal mortality of 7.5/1000 live birth and post neonatal mortality of 2.5/1000 live birth would explain the entire tric of the data!

19 March 1997

SAMPLING METHODOLOGY

By Dr. D. N. Shah

On the 3rd and last day of the workshop began with D. N. Shah narrating some common sampling methods, giving reasons for sampling and elaborating on Cluster Sampling (PPS) as it is in vogue currently and finds application in some of the routine epidemiological field work in developing countries. He mentioned in brief about the need for finding adequate sample size and the need for it to be timely and representative.

SAMPLE SIZE - WHAT SHOULD WE KNOW ?

By Maj [Dr.] A.R.N. Setalvad

The most frequently asked question of a medical statistician is "How large a sample should I take"? If it too small, we may fail to achieve the objective of our analysis; it is too big, we waste our resources and time when we gather data. This is one area where the perception of the research worker regarding the adequacy of a sample size is sharply at variance from that of a statistician. Many a times the statistician is never consulted at the planning stage or while carrying out the study; he is involved only at the stage of analysis by which time it is too late to do anything about the size of the sample.

In an ideal setup, the statistian has to be involved at the planning stage itself, particularly to determine the sample size. Although the formulae to determine the sample size are available in any standard textbook of statistics, the size of the sample is dependent on several factors, all of which contribute to its determination to a varying extent. Although the comprehensive review of all such factors is beyond the scope of this lecture, I shall make an endeavor to analyze the factors briefly.

Information Required A Priori Before Deciding Sample Size

The desirable size of a proposed study trial can be assessed using the standard formulae but before using these formulae, the information on the following points is usually required:

1. Required level of statistical significance of the expected result;
2. Acceptable chance of missing a real effect;
3. Magnitude of the effect under investigation;
4. Amount of variable present in the population.
5. Relative sizes of the groups being compared.

One point needs to be emphasized at this jucncture: the amount of information that can be gained from a sample depends upon its absolute size, not upon its size as a proportion of the population size. Some sampling error will arise because we have not studied the entire population. Whenever we sample, we miss some helpful information about the population. If we want to have a high degree of precision, we have to sample enough of the population to provide the required interpretation. Sampling error is controlled by selecting a sample that is adequate in size. In general, the more precision we want, higher is the sample size. Thus, we can use the concepts of standard error and confidence interval to help decide how many subjects should be included in the sample. Knowing something about the underlying population standard deviation is important for planning sample size. It is important that it may even be useful to coduct a pilot study to determine the population variance if no estimates are available. The flow chart for estimation of the sample size is shown in Fig. 1. However, in reality, sample size is often determined by logistic and between sample size and costs.

Magnitude Of Problem

Magnitude of problem, as defined by the levels of prevalence and incidence play a crucial role in determining the sample size. When we want to compare the differences in proportions, if the proportions themselves are small, the detection of small difference will require a large sample indeed.We can rely from the data from previous studies if they are available. However, if they are not available, it is for this reason that a pilot study to measure the prevalence / incidence [i.e. amount of variable present in the community] is absolutely essential before designing the study trial as without this information, no meaningful decision can be taken for determinig the sample size.

Representative Nature of Sample and Small Sample

There is a general feeling that a small sample is not useful for producing meaningful results. This is far from true. It is not the mere size of the sample but its representativeness which is of paramount importance. A small sample by itself does not negate the value of the study; what is important and crucial is the method of selection of sample and the true representativeness as a characteristic of the sample vis-a-vis population. Even all large sample may produce totally erroneous results if it is not truly representative while a small sample may produce fairly accurate results if it truly represents population in all respects. One of the most infamous sampling disasters [Literary Digest poll 1936 U.S.Presidency election] amply demonstrates this. In that exercise, 10 million sample ballots were mailed to prospective voters but only 2.3 million were returned. These predicted landslide victory for Ladon with 60% votes; while in fact Roosevelt won in a landslide with 62% of the vote! Although 2.3 m is a fairly large sample, it failed miserably to represent American voters in all respects with the consequential disastrous results. Statistical tests have been designed to analyze the data from small samples also provided that the sample truly represents the population. However, a small sample will also have its limitation in applying the results of the study to the population and this factor must be kept in mind while interpreting the tests of significance from small samples.

Precision Of Study

The precision of a value calculated from a sample depends upon the two factors:

(a) Size of a sample; and

(b) Variability of the characteristic within the population from which the sample is drawn.

The statistician's aim is to pass from these simple rule to more precise formulae which will enable him to estimate with a certain degree of confidence the value of mean, etc. in the population at large and also to avoid erroneous conclusions from differences between means, etc. when these differences could easily have arisen by chance. The precision of a study can also be improved by ensuring that the groups are of appropriate relative size. This is often an issue in a case-control study when a decision is required on the number of controls to be chosen for each case. Because of its relative speed and cheapness as compared with other approaches, case-control studies is a preferred method; there are difficulties in selection of cases, selection of controls and obtaining the data once the selections are made. If these precautions are not observed, case-control studies often produce contradictory and conflicting results which are wrongly attributed to small or inadequate sample size.

Systemic Error (Bias)

Although a comprehensive review of bias is beyond the scope of this lecture, a brief discussion of the bias as it affects the ultimate size of sample being analyzed is necessary here. Systemic error or bias occurs in epidemiology when there is a tendency to produce results that differ in a systemic manner from true values. A study with a smaller systemic error is said to have high accuracy; what must be borne in mind is that accuracy is not affected by sample size and is function of the observer or of the instrument. It is a particular hazard in epidemiology because we have no control over participants in such studies unlike laboratory experiments in which the control over participants or subjects is of greater magnitude. Some variables of epidemiology, particularly of subjective nature, are particularly difficult to measure [viz. Personality type, past exposure, habits, etc.] and consequently difficult to analyze. It is often difficult to obtain representative samples of source populations if such variables are to be analized. These difficulties

also lead to systemic errorr while analyzing data. The possible sources of systematic error in epidemiology are many and varied; over 30 types of such errors have been identified of which the principal ones are:

[a] Selection Bias; and [b] Measurement or Classification Bias.

Selection Bias

Select bias occurs when there is a systematic difference between the characteristics of the people selected for a study and the characteristics of those who are not. An obvious source of selection bias occurs when participants select or volunteer themselves for a study, viz. In the following situations:

1. They are unwell.
2. They are particularly worried about an exposure.
3. The perception of the respondents or the parents (in studies involving children) is different from that of the research worker.
4. Economic and social factors prevail to prevent continuous participation [for a follow-up].
5. Disease or factor itself makes people unavailable for study. This is particularly true for occupational studies where those workers who are affected are likely to leave or to be taken away from the job while those who remain are likely to be unaffected.

If the number of subjects dropping out due to reasons described above is large, the ultimate sample may be reduced to size which may not produce any valid interpretation. Another problem which often crops up while determining the size of sample is the number of control as related with the case. The first problem in a case-cotrol study is the selection of cases which usually receives little consideration beyond a definition of the type of the disease and a statement about the confirmation of the diagnosis. Howerver, the patients do not exist in isolation; the disease is the end result of a process which is almost always complex and not easy to quantify. Far greater difficult is the selection of the controls. While it is a matter of flexibility dependent upon the relative availability of cases, what is important is to ensure that there is sufficient similarity between cases and controls when the data are to be analyzed. eg. if the data were to be analyzed by age groups, the study would be inefficient leading to misleading conclusions and much of the time and effort would be wasted. While there cannot be a rule of thumb covering all situations, there is little sense in having more than four cotrols for each case; optimal figure being a matching control for each case or two controls for each case. There are two sets of controls : general population and group of patients having other diseases. Although these two population are not clearly not identical, the latter are usually preffered due to ease of availability.

The quantification of selection bias is often difficult; matching on some variables does not ensure comparability on all. If individuals entering or remaining in a study display different associations from those who do not to a major extent, a biased estimate of the association between exposure and outcome is produced.

Measurement or Classification Bias

This occurs when the measurement is inaccurate or partially accurate. There are many sources and their affects are of varying degrees. This can be offset partly by randomizing the analysis patterns; e.g. if the specimens of the exposed and cotrol groups are analyzed randomly by different laboratories with insufficient joint quality procedures, the errors will be random and potentially less harmful than in the situation wherein all the specimens from the study group are analyzed in another. If measurement bias occurs equally in the groups being compared it almost always results in an underestimate of the true strength of the relationship. This form of bias may also account for some apparent discrepancies between the results of different epidemiological studies.

Recall Bias

This is a variant of measurement bias of particular importance in retrosprective case-control studies and occurs when there is a differential recall of information by cases and controls; e.g. the cases may be more likely to recall past exposure especially if it is a well known cause and effect relationship like lack of exercise and heart disease or sickness during pregnancy and birth of a handicapped child. This can either exaggerate or underestimate the degree of effect associated with the exposure. We may be forced to rely on the unreliable memory of the subjects.

Sampling In Clinical Trials

Although randomized sample of adequate sample size is the best thing to have, most of the clinical trials are not conducted in such manner due to a variety of reasons and practical difficulties. Any study on human patients requires ethical approval and the consent of the patients which itself may be difficult to obtain. Getting the co-operation of volunteers is itself a daunting task. Therefore in general, medical studies are usually carried out using samples drawn from populations which are much more restricted than those about which we wish to draw conclusions. We may have to restrict the sample size, we may have to use patients from one hospital instead of several hospitals or all patients, or populations from a small area rather than that of a large region or state or country. We have to rely on special volunteer groups like prisoners, medical students, nurses as controls.All these factors lead to inadequate or truncated sample size which may not truly represent the entire population. Findings from such a study can only apply to the population from which the sample was drawn. Any conclusion which we come to about wider populations or countries depends upon evidence which is non-statistical and often subjective and unspecified. This may let us down circumstances, it is desirable that such studies should be repeated by other workers on other populations, so that we can sample the larger population at least to some extent by pooling the results appropriately.

Sampling In Epidemiological Studies

One of the most difficult tasks in medicine is to determine the causes of disease so that we may devise methods of prevention. However for achieving this we are working in an area where experiments are often neither possible nor ethical. We must also face the fact that disease effect and putative cause do not exist in isolation but in a complex interplay of many interrelated factors. We have to ensure that the effect which we are observing is not the result of some other factor acting on both the "Cause" and "effect". For example, it was once thought that the Yellow barked Acacia tree caused malaria because it was most frequent in those who slept under that tree! Only at a much later stage was the transmission of malaria through the bite of Anopheles was understood and it was found that this tree is a natural resting place for the mosquito as it grew close to the breeding places of the mosquito.

By nature the epidemiological designs must try to deal with the complex interrelationship between different factors in order to deduce the true mechanism of the disease. The practical difficulties which are associated with such an approach are the limitation of the sample size and the type of the study. Although cross-sectional studies are the easiest, most diseases are not suited to this simple cross-sectional approach because they are relatively rare events. We would need a very large sample indeed to get a worthwhile conclusion which is almost always impossible. Cohort studies with a prospective design is another type of study but the practical difficulties associated with them are that it takes a long period of time and involves keeping track of people over that period which itself may lead to losing the people from the sample.

Fig. 1 POINTS TO DETERMINE SAMPLE SIZE

START

To estimate a population characteristic by observing that characteristic in a sample, first make a POINT ESTIMATE:

→ Do you want to know the extent of the range of error of the estimate & the probability of the true population parameter lying within that range?

→ No → STOP

↓

YES

↓

(a) Make an interval estimate

(b) Choose Confidence Level

↓

Is the parameter

(a) Proportion ?or

(b) Mean?

↓

Is the population ?

(a) Finite or
(b) Infinite? →

(a) Normal Distribution or
(b) Other Distributions like Binomial / Poisson / Negative Binomial, etc.?

↓

Determine appropriate value of Z & SD

↓

Compute Standard Error

↓

Compute Sample size

Problems and Limitations of Epidemiology Studies :

Dr. P.V. Kotecha

The final session of the workshop on the discussion of epidemiology study was opened by Dr. Kotecha giving the summary from a study (Bulut *et al,* 1995) about the findings related to a reproductive tract infection in rural Istambul as follows :

1. Physician diagnosed infection in more than half of women who did not report the symptoms.
2. Physician diagnosed infection in only two thirds of the women who **did** report symptoms.
3. According to the laboratory examinations less than a fifth of those **reporting** a problem **had** one.
4. While 76% of those who did not recognize that they had a problem in fact had one.
5. Worse still, of the women that the physician diagnosed as not having an infection 69% actually had one.

Dr. Kotecha also described the actual data and methodology and then he invited comments from the participants. What followed in the comments and discussion was a real self-learning exercise by the participants. The doctors found faults with the laboratory and patients while the social scientists found that the point of the importance to the women's perception is the top priority.

Dr. Kotecha then explained that the expectations that the epidemiology study needs to be precise, valid and without any biases if properly planned. However in reality and in all the areas of research this may not be possible. Futher he emphasized that from what paradigm do you view is of equal importance. He then discussed the following points at length about the limitations of the epidemiology studies, which may play role many a times :

1. Over simplification of social reality
   - Preconceived categories
   - Arithmetic figures, frequencies with no real significance
   - Mystify reality
   - Photogenic (Cross-sectional)
2. Inconsistency
   - Dynamic/Genuine variability
   - Lack of truth/consistency in the replies from the respondent
   - Insensitive to the culture prevailing in the community
3. Lying Informants
   - Sensitive or personal questions where respondent does not want to reveal truth
   - Good reasons to lie with the respondent
4. Recall Problems
   - Telescopic memory
5. Missing context for explanations
   - Process
   - Linkages
   - Sequences

What followed then, was an exciting discussion about the role of medical persons in the treatment and provision of health care. Who should decide for the patients and why.

EVALUATION OF THE WORKSHOP

All the participants were asked to evaluate the workshop on an evaluation form. The following is the summary of the tabulated responses-

Majority of the participants found the management of the workshop and the quality and coverage of the reading material excellent. The quality of presentation, selection of topics, coverage of individual topics, and involvement of faculty members has been rated as good by most of the participants. The duration of the workshop was just right as per majority of the rating but some have found it longer than required and only a few found it less than required.

The aspects of training found useful by most of the participants in terms of excellence of content were the sessions - 'Limitations and problems of research studies', Epidemiological studies I', and Group work exercises for measurements in epidemiology, epidemiology studies and sampling. Most of the participants have rated the rest of the sessions as good in content with a few rating them as average. From the perspective of presentation most of the participants have rated all the sessions as good.

Asked if they would use epidemiology in the immediate future some participants would use it definitely and some hopefully in teaching, research, and P. G. guidance and the rest are either not sure or feel it is not applicable in teaching, research, P. G. guidance.

The following are the responses received from the participants in a tabulated form.

Evaluation of the workshop on

"Epidemiology: Basic Concept, Research & Women's Health"

No.	Questions	Answers Tick (√) appropriately	Freq.
1	Management of the workshop was	Excellent Good Average	27 10 00
2	Quality of presentation was	Excellent Good Average	13 23 02
3	Selection of topic was	Excellent Good Average	09 19 06
4	Coverage of individual topics was	Excellent Good Average	13 20 04
5	Involvement of faculty members was	Excellent Good Average	16 20 01

Contd...

No.	Questions	Answers Tick (√) appropriately								Freq.
6	Quality and coverage of reading material is	Excellent Good Average								21 16 01
7	Duration of the workshop was	Longer than required Just the right Less than required								10 24 04
8.	What aspects of the training you found most useful	(1. Excellent 2.Good 3.Average 4. Poor)								
			Contents				Presentation			
		1	2	3	4	1	2	3	4	
	a) Introduction to Epidemiology	15	16	06	00	14	15	02	0	
	b) Measurements & terminology used in epidemiology	06	18	12	01	08	14	09	1	
	c) Epidemiological studies -Overview	09	19	05	01	10	14	05	0	
	d) Group work for measurements in epidemiology	17	14	05	00	13	14	05	0	
	e) Epidemiological studies - I	18	15	04	00	13	14	05	0	
	f) Epidemiological studies-II	11	23	04	00	09	18	06	0	
	g) Error in epidemiological studies	09	24	02	01	09	17	03	1	
	h) Group work for epidemiology studies	17	16	01	00	05	20	02	0	
	i) Sampling methodology	01	19	17	01	02	16	11	2	
	j) Sample size : What do we need to know	07	16	11	01	07	11	10	2	
	k) Limitations & problems of research studies	20	17	02	00	16	11	02	0	
	l) Group work for Sampling	14	12	04	00	09	10	04	0	
9.	Do you think in immediate future you will start using epidemiology in near future a. Teaching	Definite Hopefully Not sure Not applicable								10 11 04 11
	b. Research	Definite Hopefully Not sure Not applicable								17 11 08 02
	c. P. G. Guidance	Definite Hopefully Not sure Not applicable								12 08 05 13

Appendix I

Program Schedule

17-3-1997

Time	Program:	Speaker
8-30 to 9-30	Registration	
9-30 to 9-40	Welcome Address	Dr. Amita Verma
9-40 to 10-00	Inauguration	Dr. R.V.Bhatt
10-00 to 10-30	About WOHTRAC Workshop Objectives	Dr. Shagufa Kapadia
10-30 to 10-45	Tea Break	
10-45 to 11-15	Introductions:	Participants
11-15 to 12-00	Introduction to Epidemiology	Dr. P. V. Kotecha
12-00 to 1-00	Measurements and Terminology Used in Epidemiology	Dr. V.K. Desai
1-00 to 2-00	Lunch	
2-00 to 3-00	Epidemiological Studies: Overview	Dr. B.S.Bhavsar
3-00 to 3-15	Discussion	
3-15 to 3-30	Tea Break	
3-30 to 4-30	Group Work: 5 Groups for Measurements in Epidemiology	
4-30 to 5-15	Presentation by the Groups	Dr. P. V. Kotecha

18-3-1997

9-30 to 10-30	Epidemiological studies: I	Dr. A. R. N. Setalvad
10-30 to 11-30	Epidemiological studies: II	Dr. Dilip Malavankar

11-30 to 11-45	Tea Break	
11-45 to 12-45	Errors in Epidemiological Studies	Dr. P. V. Kotecha
12-45 to 1-00	Discussion	
1-00 to 2-00	Lunch	
2-00 to 3-30	Group work for Epidemiology Studies	
3-30 to 3-45	Tea Break	
3-45 to 4-30	Presentation of the Group Work	
4-30 to 5-00	Discussion	S. Srinivasan

19-3-1997

9-30 to 10-30	Sampling Methodology	Dr. D. N. Shah
10-30 to 11-15	Sample Size: What do we need to know?	Dr. A. R. N. Setalvad
11-15 to 11-30	Tea Break	
11-30 to 12-30	Limitations and Problems of Research Studies	Dr. P. V. Kotecha
12-30 to 1-00	Discussion	
1-00 to 2-00	Lunch	
2-00 to 3-30	Group Work for Sampling	
3-30 to 3-45	Tea Break	
3-45 to 4-30	Presentation of the Group Work	
4-30 to 5-00	Concluding Session	Dr. G.D.Joshi

Appendix II

List of Resource Persons

Name	Position:
Dr. R. V. Bhatt	Consulting Gynecologist
Dr. D.N. Shah	Dean, Medical College, Baroda
Dr. A.N. Setalvad	Dean, Medical College, Bhavnagar
Dr. G. D. Joshi	Professor & Head, P.S.M. Department, Baroda
Dr. P. V. Kotecha	Associate Professor, P.S.M. Department, Baroda & WOHTRAC Core Group Member
Dr. D. Mavalankar	Senior Faculty, Indian Institute of Management, Ahmedabad
Dr. B. S. Bhavsar	Professor & Head, P.S.M. Department, Rajkot
Dr. S. Kapadia	Reader, H.D.F.S. Department, Faculty of Home Science& Principal Investigator, WOHTRAC
Dr. V.K. Desai	Professor & Head, P.S.M. Department, Surat
Dr. R.K. Baxi	Associate Professor, P.S.M. Department, Baroda
Dr. S. Srinivasan	Epidemiologist and Managing Trustee, Locost
Dr. Amita Verma	Director, Women's Study Research Center, Baroda & Hon. Advisor, WOHTRAC
Dr. V.S. Mazumdar	Assistant Professor P.S.M. Department, Baroda
Dr. S. Kanani	Reader, Faculty of Home Science, Baroda & Co-Investigator, WOHTRAC

Appendix III

List of Participants

Name:	Organization:
Dr. Irfan Khan	CORT
Dr. Sunanda Ganju	Deepak Foundation
Dr. Bunny Nag	Baroda Citizen Council
Dr. Aruna Lakhani	Deepak Foundation
Dr. G.K. Vankar	Professor, Psychiatry Department, Baroda
Dr. D.R. Jhala	Professor, Pediatrics Department, Baroda
Dr. Uma Nayak	Assistant Professor, Pediatrics Department, Baroda
Dr. S.R. Baxi	Assistant Professor, Pediatrics Department, Baroda
Dr. Y.S. Marfatia	Professor, Skin & V. D. Department, Baroda
Dr. Kishan Jani	Assistant Professor, Medicine Department, Baroda
Dr. Kailesh Bhalani	P.S.M. Department, Baroda
Dr. S.K. Pataudi	Ministry of Railway, Baroda
Dr. Samir Shah	P.S.M. Department, Baroda
Dr. Sangita Patel	P.S.M. Department, Baroda
Dr. J.R. Damor	P.S.M. Department, Baroda
Shalini	Faculty of Home Science, Baroda
Meena Mistry	Faculty of Home Science, Baroda
Dr. D.S. Asari	Baroda Municipal Corporation
Dr. N.R. Godara	P.S.M. Department, Baroda
Dr. Kalpana Patel	
Preeti	Faculty of Home Science, Baroda
Saroj Bhavsar	Population Research Center, Baroda
Vaishali Zararia	Baroda Citizen Council
Dr. Seema Nigam	P.S.M. Department, Baroda
Dr. Sandhya Joshi	Faculty of Social Work, Baroda
Ms. Urvi Shah	Population Research Center, Baroda
Dr. Rajaram	Faculty of Arts, Baroda
Nandini Manjrekar	Women's Study Research Center
Dr. K. Sasikala	Program Coordinator, WOHTRAC
Mausami Joshi	Research Associate, WOHTRAC
Aparna Joshi	Research Associate, WOHTRAC
Lalita Dongre	Research Associate, WOHTRAC
Dr. J.P.Mehta	P.S.M. Department, Jamnagar
Dr. Nehal Vaidya	P.S.M. Department, Rajkot
Dr. Biren Gandhi	P.S.M. Department Bhavnagar
Dr. Jasmin Shah	P.S.M. Department Bhavnagar
Dr. J.D. Lakhani	Professor & Head, Medicine Department, Karamsad
Dr. D.A. Shah	Dy. Manager (Medicals), Narmada Chematur, Bharuch
Dr. Rajesh Shah	P.S.M. Department Ahmedabad

* * * * *

Printed at Maharaja Sayajirao University of Baroda (Press) - Vadodara - 1